Abstract—We propose a bi-directional distributed antenna (DA) system where DAs are capable of working dynamically in hybrid duplex modes: full duplex (FD), half duplex (HD) and sleep, which enables higher degree of freedom and hence much higher energy efficiency (EE) than DA sole-FD systems, with just marginal loss in spectral efficiency (SE). The proposed system also demonstrates significant EE and SE enhancement over FD co-located antenna systems. Compared to the above two benchmark systems, the proposed system also requires much simpler self-interference cancellation (SIC) design for FD mode due to less cross talks between antennas. A low-complexity EE maximization scheme is proposed for the bi-directional DA system. A channel gain based DA clustering algorithm is first performed to activate/deactivate transmit/receive chains, which highlights the characteristics of DA deployment, and then a distributed hybrid duplexing (Dis-Hyb-Duplexing) algorithm is performed to optimize the downlink beamformer and the uplink transmission power. Various practical aspects are taken into account for system design, such as self-interference at DAs in FD mode, co-channel interference from uplink users to downlink users, and multiuser interference in both uplink and downlink. The effectiveness of the proposed system is verified by simulation results.

Index Terms—Hybrid duplexing, full duplex, half duplex, distributed antennas, energy efficiency

I. INTRODUCTION

Full-duplex (FD) multi-input multi-output (MIMO) communication [1] is regarded as a promising solution to meet the rapidly increasing demand for high data rate. Unlike the conventional half-duplex (HD) MIMO, where uplink and downlink transmission are decoupled into orthogonal time slots or frequencies [2] [3], a base station (BS) equipped with FD MIMO enables bi-directional communication with uplink and downlink users at the same time and frequency band [4] [5]. Due to the FD bi-directional transmission, self-interference is introduced at the BS. By self-interference cancellation (SIC) at the BS, i.e., passive suppression [6], analog cancellation and digital cancellation [7] [8], self-interference can be effectively mitigated.

Spectral efficiency (SE) maximization has been extensively researched to exploit the advantage of FD MIMO systems [5] [9] [10] [11] [12]. The authors of [5] investigated the FD transmission in various networks, such as bi-directional relay and cellular networks. The authors of [9] found that by applying mode switch, power control, interference-aware beamforming or link selection, SE of FD transmission can be improved. In [10] [11], SE of FD MIMO bi-directional systems was investigated, where small cell FD BSs were deployed to communicate with multiple users in downlink and uplink simultaneously. In [12], a multiuser FD MIMO pre-coding transceiver structure applicable for single-carrier and orthogonal frequency division multiplexing (OFDM) systems was presented to maximize SE. FD MIMO systems with heterogeneous cells were researched by [13] and [14] to improve SE. In [13], FD was applied for small cell BSs while HD was applied for macro cell BSs to relax the coverage reduction. In [14], the performance of massive MIMO wireless backhaul systems was investigated, where small cells adaptively work in FD or HD mode and each macro cell BS serves its small cells by using zero forcing beamforming. On the other hand, SE performance of FD relay systems was investigated in [15] [16]. Relay selection was proposed for a two-way FD relay system in [15], where two ends communicate with each other assisted by multiple two-way relays. A relay-assisted three-node system was investigated by [16], where the relay adaptively switches between FD and HD modes and selects the best links for transmission and reception.

Some fundamental challenges, however, need to be addressed in FD MIMO systems. The first challenge is that both MIMO and FD techniques require much higher power consumption, which is against the green evolution requirement proposed by the future communication systems. This is because a) MIMO increases the number of active transmit/receive chains, leading to much higher power consumption than single-input and single-output (SISO) [2]. b) For FD systems, additional power consumption is incurred by SIC [17] [18]. Therefore, how to ensure high energy efficiency (EE) transmission needs to be considered. However, most EE-oriented research only focuses on HD one-directional systems [19] [20] [21], which may not be directly applied to a FD bi-directional scenario due to the residual self-interference, co-channel interference from uplink users to downlink users and multiuser interference. The second challenge is with the increased number of centralized antennas, SIC circuit design.
in analog/digital domain becomes very complicated. The third challenge is the similar level of high path loss (PL) and correlated small scale fading in co-located MIMO systems. To this end, centralized FD MIMO can be formulated into a distributed manner. Several advantages can be gained: a) Since the contributions of each distributed antenna (DA) may vary practically due to the location of the users [22], system power consumption can be significantly reduced by only activating those DAs contributing the most. b) With distributed deployment, SIC becomes much easier due to the fact that the self-interference among different DAs can be well controlled in the propagation domain, benefiting from the ideal natural isolation performance. As a result, the circuit design for SIC could be as easy as the SISO case within each FD DA itself. c) The DA system can reduce the large-scale fading impact and obtain blockage-free effect using the multiple DAs distributed geographically [23], helping extend coverage and maintain connectivity of networks. In summary, there is strong evidence showing that, DA system is a remarkable alternative for co-located FD MIMO system. DA systems have been extensively utilized in practice to increase SE and extend coverage. In the USA (e.g., Michigan and Ohio), DA systems have been deployed as an important part of the landline infrastructure, as promulgated by the FCC Pole Attachment Order 11-50 [24]. DA systems have also been considered in 5G systems to form user-centric manner virtual cells [25], where users can find DAs in its vicinity to communicate with. In indoor environment, DA systems can be applied to provide seamless coverage [26], such as commercial use of WiFi. Besides, DA systems have been adopted by the International Union of Railways for high-speed rail communication systems [27] [28]. DA systems have also been investigated in academia. In [29], SE maximization and power minimization were investigated for DA systems. In [30], a fairness-aware downlink scheduling algorithm was proposed for DA systems. However, the previous research for DA systems was based on HD transmission only. In [31], FD cooperative multi-point system was studied in terms of power minimization, where the antennas can be set to either FD mode or sleep mode based on the state-of-art switch on/off technique [32]. Having only FD or sleep mode may not be optimal in some scenarios. For example, if the active antennas are only close to uplink users while are very far from downlink users, these antennas may work in HD (receive) mode to incur lower power consumption and absence of self-interference. Since HD mode is featured by lower power consumption and FD mode is featured by high SE at the cost of high power consumption, the hybrid duplexing technique enables antennas to adaptively switch among FD, HD and sleep modes, making a good trade-off between the SE and power consumption.

In this paper, we propose a bi-directional DA system where DAs are capable of working in hybrid duplex mode: FD, HD and sleep, and investigate EE maximization with dynamic hybrid duplexing. The effectiveness of the proposed system is verified by simulation. Our work is different in the following aspects:

1) Hybrid duplex mode enables higher degree of freedom and hence much higher EE compared to DA FD-sole systems [31], with just marginal loss in SE. The proposed system also demonstrates significant EE and SE enhancement over co-located FD MIMO systems [11].

2) Compared to DA sole-FD and co-located FD MIMO systems, the proposed DA hybrid duplexing system requires much simpler SIC design for FD mode, due to less cross talks between DAs.

3) We investigate EE maximization by considering both bi-directional SE and power consumption. This is different from the existing work on SE maximization alone [9]-[11] or total power consumption minimization alone [31]. Various practical aspects are taken into account for the optimization, such as residual self-interference at DAs in FD mode, co-channel interference from uplink users to downlink users, and multiuser interference in both uplink and downlink, whereas only self-interference was considered in our previous work on one-directional FD relay systems [4] [18]. Also, the power amplifier (PA) dissipated power, circuit power, system fixed power and SIC power are all included in the power consumption model, which is more accurate than the existing models formulated for FD systems [31] [33].

4) Low-complexity algorithms are proposed for EE maximization. A channel gain based DA clustering algorithm is first performed to activate/deactivate transmit/receive chains, which highlights the characteristics of the distributed deployment, and a distributed hybrid duplexing (Dis-Hyb-Duplexing) algorithm is then performed to optimize the downlink beamformer and the uplink transmission power.

**Notations:** Matrices and vectors are represented by boldface capital and lower case letters, respectively. $|·|$ denotes the absolute value of a complex scalar. $\|·\|$ denotes the Euclidean vector norm. $A^H$ and $\text{Tr}(A)$ denote the Hermitian transpose and trace of matrix $A$. $\text{Rank}(A)$ denote the rank of matrix $A$. $\text{diag}(A)$ returns a diagonal matrix with diagonal elements from matrix $A$. $A \succeq 0$ means $A$ is a positive semi-definite matrix. Superscript $k$ or $u$ denote the downlink/uplink users’ indexes.

## II. System Model and Problem Formulation

### A. System Model

We consider a bi-directional DA system with one central signal processing unit, which consists of a baseband module and $L$ DAs. All DA ports are connected to the central unit through a noise-free wired front-haul for cooperative communications. The DAs can work in FD, HD (transmit or receive) or sleep mode, while the users employ single-antenna HD for low hardware complexity. In particular, there are $U$ uplink users and $K$ downlink users. Channel state information (CSI) is obtained by channel estimation in the training phase for both uplink and downlink, based on channel reciprocity as in bi-directional FD MIMO systems [34] [35] [36]. We assume that all the downlink/uplink users and DAs utilize orthogonal pilot sequences, which allows the DAs to distinguish the pilots of different users. In the training phase, the uplink users send pilots to the DAs and uplink CSI can be estimated at the DAs. The downlink CSI can be obtained through the channel quality indicator from downlink users [34]. Hence, the DAs
can regularly update the CSI estimates of the uplink and downlink channels [34] [36]. Also, each DA can send pilots to other DAs, and thus self-interference matrix can be obtained at all DAs. For the co-channel interference channels, since the downlink users can receive the pilots from the uplink users and feed back the estimates for DAs, the DAs can update the CSI of the co-channel interference channels as well [34]. The system model is depicted by Fig. 1.

Define a receive chain activation vector \( r \in \mathbb{C}^{L \times 1} \), whose \( l \)-th element \( r_l \) equals to 1 if the \( l \)-th DA's receive chain is activated, otherwise \( r_l = 0 \). Define a transmit chain activation vector \( t \in \mathbb{C}^{L \times 1} \), whose \( l \)-th element \( t_l \) equals to 1 if the \( l \)-th DA's transmit chain is activated, otherwise \( t_l = 0 \). The configuration is reasonable in practice, since the shared-antenna deployment has been extensively researched, where only one antenna set is adopted for simultaneous transmission and reception, and the received and transmitted signals are separated with the help of a duplexer [37] [38]. Define \( w_k \in \mathbb{C}^{L \times 1} \) as beamformer at DAs for downlink user \( \forall k \in K \). For each vector \( w_k \), its element \( w^l_k \) denotes the beamforming weight of DA \( l \) for user \( k \). Denote \( p_u \) as transmission power allocation at uplink users, for \( \forall u \in U \).

#### B. Problem Formulation

To maximize the system EE, we jointly optimize DA vectors \( r, t \), beamformer \( w_k \), for \( \forall k \in K \), and transmission power \( p_u \), \( \forall u \in U \). Define \( \omega(r, t, w_k, p_u) \) as the EE (in bits/Joule/Hz), which is the ratio of the system total throughput \( T_{total} \) to the incurred total power consumption \( P_{total} \). Accordingly, the EE maximization problem is formulated as

\[
P_1 : \quad \text{argmax}_{r, t, w_k, p_u, \forall k \in K, \forall u \in U, \sum l} \frac{T_{total}}{P_{total}},
\]

subject to:

\[
(C1) : \quad 0 \leq \sum_{k=1}^{K} |w^l_k|^2 \leq p_{DA}, \forall l \in L,
\]

\[
(C2) : \quad 0 \leq p_u \leq p_{u,max}, \forall u \in U,
\]

\[
(C3) : \quad r_l = \{0, 1\}, \forall l \in L,
\]

\[
(C4) : \quad t_l = \{0, 1\}, \forall l \in L,
\]

where \((C1)\) denotes that the allocated transmission power at each DA is non-negative and upper bounded by \( p_{DA} \), for \( \forall l \in L \). \((C2)\) denotes that the transmission power allocated at each uplink user \( u \) is non-negative and upper bounded by \( p_{u,max} \), for \( \forall u \in U \). \((C3)\) and \((C4)\) denote the statuses of the receive and transmit chains of DA \( l \), for \( \forall l \in L \).

### III. THROUGHPUT AND POWER CONSUMPTION ANALYSIS

The optimization problem involves with the throughput and power consumption in both downlink and uplink. In this section, we first analyze the downlink and uplink throughput in Subsection III-A and III-B, respectively, and then give the total power consumption in Subsection III-C.

#### A. Downlink Throughput

In each time slot, the received signal at downlink user \( k \) is given by

\[
y_{k}^{DL} = h_k(t \circ w_k)d_{k}^{DL} + \sum_{k' \neq k} h_{k}(t \circ w_{k'})d_{k'}^{DL} + \sum_{l=1}^{U} \sum_{u=1}^{U} \sqrt{p_u} c_{u,k} d_{u,l}^{UL} + z_k,
\]

where the operator \( \circ \) denotes Hadamard product of two vectors or matrices. Vector \( h_k \in \mathbb{C}^{L \times 1} \) is downlink channel between \( L \) DAs and downlink user \( k \) and captures the PL and small scale fading, whose \( l \)-th element \( (h_k)_l \) presents the channel condition between the \( l \)-th DA and downlink user \( k \). \( d_{k}^{DL} \in \mathbb{C} \) is transmitted data for downlink user \( k \), \( e_{u,k} \in \mathbb{C} \) denotes the channel condition from uplink user \( u \) to downlink user \( k \). \( d_{u,l}^{UL} \in \mathbb{C} \) is the transmitted data from uplink user \( u \). \( z_k \sim CN(0, \sigma^2_k) \) is the complex Additive White Gaussian Noise (AWGN) at downlink user \( k \). Without loss of generality, we assume \( E\{(d_{k}^{DL})^2\} = E\{(d_{u,l}^{UL})^2\} = 1 \), for \( \forall k \in K \) and \( \forall u \in U \). For simplicity, we introduce the auxiliary matrices \( F_l = \text{diag}\{0, ..., 0, 1, 0, ..., 0\} \) to transform the Hadamard product \( t \circ w_k \) into a simpler form \( t \circ w_k = \sum_{l=1}^{L} t_l F_l w_k \). Therefore, equation \((2)\) can be re-expressed as

\[
y_{k}^{DL} = h_k(\sum_{l=1}^{L} t_l F_l w_k)d_{k}^{DL} + \sum_{k' \neq k} h_{k}(\sum_{l=1}^{L} t_l F_l w_{k'})d_{k'}^{DL} + \sum_{u=1}^{U} \sum_{u=1}^{U} \sqrt{p_u} e_{u,k} d_{u,l}^{UL} + z_k.
\]

According to \((3)\), the signal-to-interference-and-noise ratio (SINR) of user \( k \) is calculated as

\[
\Gamma_{k}^{DL} = \frac{|h_k(\sum_{l=1}^{L} t_l F_l w_k)|^2}{\sum_{k' \neq k} |h_k(\sum_{l=1}^{L} t_l F_l w_{k'})|^2 + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma^2_k}.
\]
Thus, the total downlink throughput is calculated as

\[ T_{DL} = \sum_{k=1}^{K} \log_2(1 + G_{DL}^k). \]  

(5)

Substituting (4) in (5) and transforming (5) into the structure of \( \log(1 + \frac{a}{b}) = \log(A + B) - \log B \), we have

\[ T_{DL} = \sum_{k=1}^{K} \log_2 \left( \sum_{k'=1}^{K} \sum_{l=1}^{L} |h_{k'}(l)|^2 + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2 \right) - \sum_{k=1}^{K} \log_2 \left( \sum_{k' \neq k}^{K} \sum_{l=1}^{L} |h_{k'}(l)|^2 + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2 \right). \]  

(6)

Defining matrices \( H_k = h_k^H h_k \) and \( W_k = w_k w_k^H \) and substituting them into (6), the total downlink throughput is finally given by

\[ T_{DL} = \sum_{k=1}^{K} \log_2 \left( \sum_{k'=1}^{K} \text{Tr} \left( \sum_{m=1}^{L} \sum_{u=1}^{U} t_m t_u F_m W_{k'} F_n^H H_k \right) \right) + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2 \]  

\[ - \sum_{k=1}^{K} \log_2 \left( \sum_{k' \neq k}^{K} \text{Tr} \left( \sum_{m=1}^{L} \sum_{u=1}^{U} t_m t_u F_m W_k F_n^H H_k \right) \right) + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2 \). \]  

(7)

### B. Uplink Throughput

In uplink, define self-interference channel matrix \( H_{SI} \in \mathbb{C}^{L \times L} \), whose element \((H_{SI})_{ij}\) represents the channel condition from the \( j \)-th DA to the \( i \)-th DA. Specifically, the diagonal entry \((H_{SI})_{ii}\), for \( i \in L \), represents the \( i \)-th DA’s self-interference channel. The self-interference signal is the downlink transmitted signal from the DAs, and thus can be calculated as \( \sum_{k=1}^{K} \sum_{l=1}^{L} t_l F_l w_k d_{DL}^L \). Define an SIC amount \( \alpha \) as the ratio of post-SIC self-interference power over the pre-SIC self-interference power. The residual self-interference after SIC is thus given by \( \frac{1}{\sqrt{\alpha}} H_{SI} \sum_{k=1}^{K} \sum_{l=1}^{L} t_l F_l w_k d_{DL}^L \), and the uplink received signal at the DAs is given by

\[ y_{UL}^{UL} = \sum_{u=1}^{U} \sqrt{p_u} g_u d_{UL}^u + \frac{1}{\sqrt{\alpha}} H_{SI} \sum_{k=1}^{K} \sum_{l=1}^{L} t_l F_l w_k d_{DL}^L + z, \]  

where vector \( g_u \in \mathbb{C}^{L \times 1} \) is the uplink channel from user \( u \) to \( L \) DAs, whose element \((g_u)_i\) presents the channel condition between the \( i \)-th DA and uplink user \( u \). Vector \( z \in \mathbb{C}^{L \times 1} \sim \mathcal{CN}(0, \sigma^2 I_L) \) is the AWGN receive noise. We assume that the central unit employs a linear maximum ratio combining (MRC) receiver [31], \( v_u \in \mathbb{C}^{L \times 1} \), for decoding of the received \( u \)-th uplink user information, which is given by \( v_u = \sum_{i=1}^{L} r_i F_i g_u \). It is obvious that \( r_i = 0 \) means the receive chain of the \( i \)-th DA is deactivated. As a result, the equivalent SINR for uplink user \( u \) is given by

\[ \Gamma^U_u = \frac{p_u |v_u^H g_u|^2}{\sum_{u' \neq u}^U p_{u'} |v_{u'}^H g_u|^2 + \sum_{u=1}^U |v_u^H H_{SI} \sum_{k=1}^{K} \sum_{l=1}^{L} t_l F_l w_k|^2 + ||v_u||^2 \sigma^2}. \]  

Therefore, the total uplink throughput is given by

\[ T_{UL} = \sum_{u=1}^{U} \log_2(1 + \Gamma^U_u). \]  

(10)

Substituting (9) into (10) leads

\[ T_{UL} = \sum_{u=1}^{U} \log_2 \left( \sum_{u'=1}^{U} p_{u'} |v_{u'}^H g_u|^2 + \frac{|v_u^H H_{SI} \sum_{k=1}^{K} \sum_{l=1}^{L} t_l F_l w_k|^2}{\alpha} + ||v_u||^2 \sigma^2 \right) \]  

\[ - \sum_{u=1}^{U} \log_2 \left( \sum_{u' \neq u}^U p_{u'} |v_{u'}^H g_u|^2 + \frac{|v_u^H H_{SI} \sum_{k=1}^{K} \sum_{l=1}^{L} t_l F_l w_k|^2}{\alpha} + ||v_u||^2 \sigma^2 \right). \]  

(11)

Substituting \( G_u = g_u g_u^H \) and \( v_u = \sum_{i=1}^{L} r_i F_i g_u \) into (11), the total uplink throughput is given by (12). Removing the quadratic terms in (12), (12) is finally derived into (13).

### C. Power Consumption

As researched by [39] [40] [41], the power consumption in DA systems mainly includes PA power consumption, circuit power consumption and system fixed power consumption [19] [23] [42]. Especially with FD technique, additional power may be incurred for SIC operation.

a) It was pointed out in the FP7 EARTH project [43] that for a small-scale communication node (such DA, femto or pico node), its PA power accounts for 40%-47% of the total power [43]. The PA power is closely related to the beamforming weights at DAs, uplink transmission power at users and drain efficiency (DE) of PA, which is given by \( \frac{1}{2} (\text{Tr}(\sum_{k=1}^{K} W_k) + \sum_{u=1}^{U} p_u) \) and \( \eta \) is the DE of PAs. Without loss of generality, we assume that all PAs (of DAs and users) work in linear region and have the same DE performance [23]. Besides, per-DA and per-uplink user power constraints are applied. This is because the DAs and users have individual transmit chain and cannot share power with each other.

\(^1\)It is known that MRC provides a performance close to minimum mean square error (MMSE) beamforming given a large number of antennas [44]. Besides, since the coupling between \( r, t, W_k \) and the residual self-interference at DAs, it is difficult to derive a tractable solution by applying MMSE or zero-forcing receive beamforming.
\[ T_{UL} = \sum_{u=1}^{U} \log_2 \left( \sum_{u'=1}^{U} p_{u'} \right) \left( \sum_{i=1}^{L} r_i F_i g_{u'}^H g_{u'} \right)^2 + \frac{1}{\alpha} \left( \sum_{k=1}^{K} \sum_{j=1}^{L} t_{k} W_{k} W_{k}^H \right) \sum_{i=1}^{L} t_{i}^2 f_i w_k \right)^2 + \left\| \left( \sum_{i=1}^{L} r_i F_i g_{u'}^H \right) \|^2 \sigma^2 \right) \] (12)

\[
T_{UL} = \sum_{u=1}^{U} \log_2 \left( \sum_{u'=1}^{U} p_{u'} \right) \left( \sum_{i=1}^{L} r_i F_i g_{u'}^H g_{u'} \right)^2 + \frac{1}{\alpha} \left( \sum_{k=1}^{K} \sum_{j=1}^{L} t_{k} W_{k} W_{k}^H \right) \sum_{i=1}^{L} t_{i}^2 f_i w_k \right)^2 + \left\| \left( \sum_{i=1}^{L} r_i F_i g_{u'}^H \right) \|^2 \sigma^2 \right) \] (13)

b) Circuit power consumption is proportional to the number of active transmit and receive chains [39]. Transmit chain and receive chains contain multiple components, such as digital/analog (D/A) converter, analog/digital (A/D) converter, electrical/photonic (E/O) converter, optical/electrical (O/E) converter, up-converter, down-converter, filter, synthesizer, mixers, etc [23]. Therefore, the power consumption of transmit chain and receive chains can be formulated by \( p_{r,t} \) and \( p_{c,t} \), respectively. When both transmit and receive chains are deactivated, the DA is in sleep mode and the power consumption is greatly reduced to \( p_{idle} \) based on the state-of-the-art switch-off technique [31]. Therefore, the total circuit power consumption is calculated as \( \sum_{l=1}^{L} \left( (1-r_l)(1-t_l) p_{idle} \right) + \sum_{l=1}^{L} (r_l p_{c,r}) + \sum_{l=1}^{L} (t_l p_{c,t}) \).

c) System fixed power consumption \( p_{fix} \) is the power consumed by power supply, active cooling system at central unit and/or DAs, etc., which is independent of the state of transmit/receive chain [31].

d) For the DAs working in FD mode, i.e., \( r_l = 1 \) and \( t_l = 1 \), for \( \forall l \in L \), additional power is required for SIC, which is related to the specific SIC circuit design \(^2\), and can be considered as a constant value [45] [46]. It is because passive suppression benefits from PL and antenna isolation, which actually do not consume additional power. While the power of analog and digital cancellation is non-negligible compared to passive suppression. The power consumed by the involved components, such as filter, attenuator, splitter and adder can be formulated as constant [18]. Define \( p_{can} \) as the power required by SIC at each DA, and the overall power consumed by SIC is given as \( \sum_{l=1}^{L} (r_l p_{can}) \).

Finally, the total power consumption is given by

\[
P_{total} = \frac{1}{\eta} \left( \sum_{k=1}^{K} W_{k} \right) + \sum_{u=1}^{U} p_{u} + \sum_{l=1}^{L} (t_l p_{can}) + p_{fix} + \sum_{l=1}^{L} \left( (1-r_l)(1-t_l) p_{idle} \right) + \sum_{l=1}^{L} (r_l p_{c,r}) + \sum_{l=1}^{L} (t_l p_{c,t}) \].

IV. ENERGY EFFICIENCY (EE) PERFORMANCE OPTIMIZATION

According to the throughput and power consumption analyzed in Section III, the problem P1 is re-formulated into

\[
P2: \quad \text{argmax}_{r,t,W_k,p_u} \quad T_{DL}(r,t,W_k,p_u) + T_{UL}(r,t,W_k,p_u),
\]
\[ \text{s.t.} \quad (C1) : \sum_{k=1}^{K} \text{Tr}(W_k F_i) \leq p_{DA}, \forall l \in L, \]
\[ (C2), (C3), (C4), \]
\[ (C5) : W_k \succeq 0, \forall k \in K, \]
\[ (C6) : \text{Rank}(W_k) = 1, \forall k \in K, \]

\(^2\)Generally speaking, complex SIC scheme needs more involved components and consumes higher power. For example, the analog cancellation design in [1] needs D/A converter, transmit radio unit and adder to mitigate self-interference. Another kind of design in [17] uses tunable attenuator and delay unit to route the estimated signal to the receiver for SIC.

\(^3\)Herein, analog or digital of SIC is only required within each FD DA itself, which is because the self-interference among different DAs can be well controlled in the propagation domain benefiting from the high PL. Also, SIC operation across different DAs imposes high complexity for hardware design. More details can be referred in [1] for co-located FD MIMO systems.
where constraint \((\tilde{C}1)\) denotes that the transmission power at each DA is upper bounded by \(p_{DA}\), for \(\forall l \in L\), (C5) and (C6) are imposed to guarantee that \(W_k = u_ku_k^H\) holds after optimization, for \(\forall k \in K\).

The problem in P2 is generally very difficult to solve. It involves a) vectors \(r, t\) with binary elements in the objective function and the combinatorial constraints in (C3), (C4). b) product of binary variables \(t_m, t_n, r_i, r_j\) with continuous variables \(W_k, p_u\), for \(\forall k \in K, u \in U\) and \(\forall m, n, i, j \in L\). c) rank constraint in (C6). Besides, with \(L\) DAs, there are \(4^L\) possibilities of DA configurations. It is prohibitive to find the global optimum in terms of computational complexity and thus a low-complexity suboptimal design is desirable. In the Subsection IV-A, we first propose a channel-gain-based DA clustering algorithm to perform DAs activation/deactivation, Subsection IV-A, we first propose a channel-gain-based DA clustering algorithm to perform DAs activation/deactivation, which effectively removes the binary variables in the objective function and the constraints (C3) (C4), as well as the product between binary variables with continuous variables. After this, a novel distributed hybrid duplexing (Dis-Hyb-Duplexing) algorithm is proposed to optimize beamformer \(W_k\), for \(\forall k \in K\) and transmission power \(p_u\), for \(\forall u \in U\), in Subsection IV-B. At last, total complexity is given in IV-C.

A. Activation/Deactivation of Transmit/Receive Chains at DAs

Intuitively, if the channel gain between the \(l\)-th DA and uplink users is strong, while the channel gain between the \(l\)-th DA and downlink users is poor, the \(l\)-th DA should work in HD receive mode. Since if the DA works in FD mode, its downlink throughput contribution is marginal for all downlink users, while more power consumption is required. More importantly, the downlink transmission corrupts its uplink reception due to the introduced self-interference. Conversely, the DA should work in HD transmit mode, if the channel gain between the DA and uplink users is poor while the channel gain between the DA and downlink users is strong. Also, one DA can work in FD mode if it has good channel condition in both uplink and downlink, which contributes reasonable throughput bidirectionally with affordable power consumption. At last, the DA can be turned off to save power consumption if it has poor channel condition from all users.

To implement the DA clustering algorithm, a threshold, i.e., \(\psi\), is needed for judgment. The threshold can be adjusted according to different quality of service (QoS) requirements and densities of DAs. For example, one can increase \(\psi\) if power consumption requirement is stringent. It means less DAs will be activated and therefore power consumption is decreased. If users’ SE requirement is stringent, on the other hand, one can decrease the value of \(\psi\), which means more DAs will be activated. Also, for a dense DA deployment, the value of \(\psi\) could be higher than a sparse deployment. The whole DA clustering algorithm is presented in Algorithm 1, then downlink beamformer and uplink transmission power allocation are ready to perform.

B. Design of Downlink Beamformer at DAs and Uplink Transmission Power at Users

For simplicity, we define a super matrix \(W = \{W_1, W_2, ..., W_K\}\), including all beamformer variables \(W_k, \forall k \in K\). Define a vector \(p = \{p_1, p_2, ..., p_U\}\), including all uplink transmission power variables \(p_u, \forall u \in U\). The feasible domain confined by the constraints is expressed as \(\Theta\). Then the optimization problem goes into

\[
P3 : \max_{W, p} \frac{T_{DL}(W, p) + T_{UL}(W, p)}{P_{total}(W, p)},
\]

s.t. \((\tilde{C}1), (C2), (C5)\) and \((C6)\).

For the total throughput \(T_{DL}(W, p) + T_{UL}(W, p)\), we collect the positive parts into \(f_1(W, p)\), and the negative parts into \(f_2(W, p)\). Then the total throughput can be expressed as \(f_1(W, p) - f_2(W, p)\), where \(f_1(W, p)\) and \(f_2(W, p)\) are given by (17) and (18), respectively.

Obviously, \(f_1(W, p)\) and \(f_2(W, p)\) are both jointly-concave with respect to the variables \(W, p\) in the considered domain, and the Frank-Wolfe (FW) method can be adopted to handle the difference of two concave functions. The FW method approximates the minus part \(f_2(W, p)\) by its first order Taylor series \(f_2^{(n)}(W, p)\), and updates the approximation \(f_2^{(n)}(W, p)\) iteratively along the direction that approaches the original function \(f_2(W, p)\) [11]. Suppose the value of \(W, p\) is denoted by \(W^{(n)}, p^{(n)}\) and \(f_2(W, p)\) is approximated by \(f_2^{(n)}(W, p)\) at the \(n\)-th iteration. Since the total throughput \(f_1(W, p) - f_2^{(n)}(W, p)\) is the lower bound of the original one \(f_1(W, p) - f_2(W, p)\), the variables \(W^{(n)}, p^{(n)}\) are iteratively updated and the lower bound increases after every iteration. Because of the power consumption is upper bounded by the constraints, the iterative procedure is guaranteed to converge.

Since \(f_2(W, p)\) is concave and differentiable on the considered domain, one can easily find its first-order approximation, as

---

**Algorithm 1 DA clustering algorithm**

**Input:** \(h_k, g_u\) for \(\forall k \in K, u \in U\), and the threshold \(\psi\).

**Output:** Transmit/receive chains vector \(r^*, t^*\).

1. For \(l = 1, ..., L\) do
2. For \(u = 1, ..., U\) do
3. Calculate the channel condition between the \(l\)-th DA and uplink user \(u\).
4. If \((g_u)_l \geq \psi\) then
   - \(r_l = 1\), break.
5. Else
   - \(r_l = 0\).
6. End if
7. End for
8. End for
9. For \(k = 1, ..., K\) do
10. Calculate the channel condition between the \(l\)-th DA and downlink user \(k\).
11. If \((h_k)_l \geq \psi\) then
12. \(t_l = 1\), break.
13. Else
14. \(t_l = 0\).
15. End if
16. End for
17. End for
18. End for
19. Return \(r^*, t^*\).
shown by (19) on the next page. The related variables \((a_k^{(n)})\) and \((b_u^{(n)})\) in (19) are detailed by (20) and (21), respectively.

Now, the overall throughput is re-expressed as \(f_1(W, p) - f_2^{(n)}(W, p)\), which is jointly-concave with respect to the variables. It is because \(f_1(W, p)\) is jointly-concave, and \(f_2^{(n)}(W, p)\) is affine with respect to all variables \(W, p \in \{\Theta\}\). The optimization problem can be expressed as

\[
P4: \arg\max_{W, p \in \{\Theta\}} \frac{f_1(W, p) - f_2^{(n)}(W, p)}{P_{\text{total}}(W, p)},
\]

s.t. \((C1), (C2), (C5)\) and \((C6)\).

Since the problem \(P4\) is the ratio between a concave function and an affine function in the consider domain, we introduce Theorem 1 to solve the problem.

**Theorem 1:** The reformulated EE, \(\frac{f_1(W, p) - f_2^{(n)}(W, p)}{P_{\text{total}}(W, p)}\), is jointly quasi-concave with respect to variables \(W\) and \(p\) in the feasible domain.

**Proof:** See APPENDIX A.

Theorem 1 confirms that there is a global optimal EE in the feasible domain. However, the objective function of (22) is in a fractional structure, which is still difficult to handle. Therefore, we further introduce Proposition 1 to transfer (22) into an equivalent but easier structure.

**Proposition 1:** Assume that \(\beta^*\) is the maximum value of \(\frac{f_1(W, p) - f_2^{(n)}(W, p)}{P_{\text{total}}(W, p)}\) in (22). Maximizing the quasi-concave problem in (22) is equivalent to finding the root of a subtract programming problem \(f_1(W, p) - f_2^{(n)}(W, p) - \beta^* P_{\text{total}}(W, p)\) [47] [48].

**Proof:** See APPENDIX B.

According Proposition 1, the problem in (22) can be associated with an equivalent subtract programming problem. Thus solving the problem \(P4\) turns into solving \(f_1(W, p) -
\( f_2(n)(W, p) - \beta P_{\text{total}}(W, p) \) with a given \( \beta \), as shown by the problem P5 in (23).

\[
\text{P5 : } \arg\max_{W, p \in \Theta} f_1(W, p) - f_2(n)(W, p) - \beta P_{\text{total}}(W, p), \quad \text{s.t. } (C1), (C2), (C5), (C6).
\]

Now we introduce Proposition 2 to handle the rank one constraint in (C6).

**Proposition 2**: For the rank constraint in (C6), rank relaxation can be applied [34] [47] [49]. In particular, with (fractional-structure)

\begin{align*}
(\hat{b}_u(n)) = \sigma^2 \text{Tr} \left( \sum_{i=1}^{L} \sum_{j=1}^{L} r_i r_j F_i G_u F_j^H \right) + \sum_{u' \neq u} \left( \rho_{u'}(n) \text{Tr} \left( \sum_{i=1}^{L} \sum_{j=1}^{L} r_i r_j F_i G_u F_j^H \right) \right) + \\
\frac{1}{\alpha} \text{Tr} \left( \text{diag} \left( H_{ST} \sum_{k=1}^{K} \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n F_m W_k^{(n)} F_n^H \right) H_{ST}^H \right) \left( \sum_{i=1}^{L} \sum_{j=1}^{L} r_i r_j F_i G_u F_j^H \right).
\end{align*}

(21)

Algorithm 2 Dis-Hyb-Duplexing algorithm

**Input**: Antenna configuration vectors \( r, t \), left/right bounds \( \beta_1 \) and \( \beta_r \), channel condition, i.e., \( h_k, g_u, e_{u,k}, h_{SI} \), for \( \forall k \in L, u \in U \), and power consumption parameters, i.e., \( \eta, p_{idle}, p_{ct}, p_{ct}, p_{bk}, p_{tx}, p_{DA}, p_{u,\text{max}} \).

**Output**: Optimal beamforming weight \( W^*_k \), for \( \forall k \in L \), and optimal uplink transmission power \( p_u^*, \forall u \in U \).

1. Set accuracy factor \( \epsilon > 0 \), and suppose \( F(\beta) \) is the optimal value of \( f_1(W, p) + f_2(n)(W, p) - \beta P_{\text{total}}(W, p) \).
2. Initialize left bound \( \beta_l \) and right bound \( \beta_r \) that ensure the \( F(\beta) \cdot F(\beta_r) < 0 \).
3. while \( \beta_r - \beta_l > \epsilon \) do
4. Solve the problem P5 with the FW method until convergence.
5. if \( F(\beta_l) \cdot F(\beta) < 0 \) then
6. \( \beta_r = \beta_l \).
7. else
8. \( \beta_l = \beta_r \).
9. end if
10. end while

selects a subinterval in which a root must lie for further processing. The bisection method is guaranteed to converge to a root of a function if the function is continuous between the left and right bounds, and the values of the functions have opposite signs at the two bounds.

Assume that \( \beta_1 = \frac{\beta_l + \beta_r}{2} \) is the midpoint of the initial interval, and \( \beta_n \) is the midpoint of the interval in the \( n \)-th step. Then the difference between \( \beta_n \) and \( \beta^* \) is bounded by \( |\beta_n - \beta^*| \leq \frac{\beta_r - \beta_l}{2^n} \). With a tolerance factor \( \epsilon \), the required number of iterations is given by \( n \leq \log_2 \left( \frac{\beta_r - \beta_l}{\epsilon} \right) \). In the proposed algorithm, the left bound \( \beta_l \) can be set to 0, with which the value of equation \( F(\beta_l) \) is definitely positive. Also, a sufficiently large value of \( \beta_f \) can be chosen as the right bound to make the value of \( F(\beta_f) \) negative. Therefore, function \( F(\beta) \) has opposite signs at the two bounds and thus the bisection approach readily leads to convergence. In the inner layer, CVX is called to solve the SDP optimization, whose complexity is indexed as \( \xi \). Thus the complexity of the Dis-Hyb-Duplexing algorithm is \( O(\log_2 \left( \frac{\beta_r - \beta_l}{\epsilon} \right) \cdot \xi) \). As a result, the total complexity is given by \( O(L(K + U) + \log_2 \left( \frac{\beta_r - \beta_l}{\epsilon} \right) \cdot \xi) \).

As can be seen, the associated bisection search in the outer

\footnote{Since the FW method is adopted for the inner layer iteration, the convergence rate is \( O(\frac{1}{L}) \) [50]. In the \( n \)-th update in the inner layer, the CVX SDPT3 solver implements an interior-point method to solve the SDP problem, which belongs to the class of path-following method and leads a fast convergence.}
layer has the largest impact on the complexity of the whole algorithm, more insight on the number of iterations of the associated bisection search will be demonstrated in Section V.

V. SIMULATION RESULTS

We use numerical results to verify our analysis, with 4000 times of Monte Carlo simulations run to obtain an average and common parameters given in TABLE I. The SIC amount is set to $\alpha = 100$ dB except for Fig. 4, where $\alpha$ varies from 60 to 100 dB. The DA clustering threshold is set to $\psi = 1 \times 10^{-9}$ except for Fig. 6, where $\psi$ varies from $0.5 \times 10^{-9}$ to $5 \times 10^{-9}$. The power constraint is set to $p_{DA} = 100$ mW except for Fig. 5, where $p_{DA}$ varies from 60 mW to 140 mW. The PL model of $PL(d) = 145.4 + 37.5 \log_{10}(d/1000)$ [51] is adopted, which is featured in 3GPP LTE standards at 2 GHz. The small-scale fading is modeled as Rician fading, including the self-interference channel $H_{SI}$, with Rician factor 5 dB [52]. The DAs are uniformly deployed across an area of 100 m $\times$ 100 m, with an example of 25 DAs demonstrated in Fig. 2. 25 DAs are used for all figures except Figs. 6 and 7, where the number of DAs varies from 16 to 36. 2 uplink users with locations (92 m, 10 m) and (85 m, 86 m), respectively, and 2 downlink users with locations (5 m, 76 m) and (78 m, 16 m), respectively, are used for Figs. 3-7. While in Figs. 6 and 9, the total number of users is increased from 4 to up to 16, and their locations are independent and identically distributed (i.i.d.) random variables following uniform distribution.

The distributed sole-FD system [31] and the co-located FD MIMO system [47] are used as benchmarks. In the former, all antennas are distributed but are only capable of working in FD or sleep modes. In the latter, all antennas are centralized at the BS. Since all antennas have identical PL to users due to centralized antenna deployment, the activation/deactivation of transmit/receive chains is disabled. Hence, all antennas work in FD mode and all elements in the self-interference channel matrix $H_{SI}$ become positive.

Fig. 3 shows the convergence behavior of the proposed Dis-Hyb-Duplexing algorithm on finding the optimal value of $\beta^*$. With left bound $\beta_l = 0$, right bound $\beta_r = 100$ and tolerance $\epsilon = 0.1$, at most 10-12 iterations are required to achieve convergence, which matches the analysis of $n \leq \log_2(\frac{\beta_r-\beta_l}{\beta^*})$ iterations in Subsection IV-C and confirms the low complexity of the proposed algorithm.

Fig. 4(a) demonstrates the average optimal EE performance under three systems. The centralized antennas of co-located FD MIMO system are located in the center of the map. It can be seen that the proposed system outperforms others in terms of EE at all SIC settings. It is because under the proposed system, the DAs with poor channel condition are deactivated, hence power waste is avoided. Besides, the activated DAs can work in HD modes (only transmit or receive) if the DAs have only reasonable channel gain in one direction. Differently, under the distributed sole-FD system, all activated DAs work in FD mode. This leads to much higher power consumption than the proposed system, but not all DAs contribute acceptable throughput in bi-direction, hence the EE is degraded. For the co-located FD MIMO system, all the centralized antennas suffer similar PL, leading to poor throughput performance. Besides, high power consumption is incurred due to the fully activated antennas. As a result, the EE performance is the lowest among the three systems. Also, it can be seen that, for all the three systems, EE

<table>
<thead>
<tr>
<th>TABLE I. Simulation Setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central carrier frequency</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>AWGN power spectral density</td>
</tr>
<tr>
<td>Cell model</td>
</tr>
<tr>
<td>DE of PA $\eta$</td>
</tr>
<tr>
<td>$p_{idle}$, $p_{c,\nu}$ and $p_{c,v}$</td>
</tr>
<tr>
<td>$p_{can}$ and $p_{f,\nu}$</td>
</tr>
<tr>
<td>$p_{u,max}$</td>
</tr>
<tr>
<td>Antenna Gain</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
</tbody>
</table>

![Fig. 2](image2.png) 25 DAs are uniformly fixed across the area. Without loss of generality, the DAs are indexed by column.

![Fig. 3](image3.png) The convergence behavior on finding optimal $\beta^*$ (equals to optimal EE) by bisection search, with left bound $\beta_l = 0$, right bound $\beta_r = 100$ and tolerance $\epsilon = 0.1$. |
will be improved with a increased SIC amount due to the less residual self-interference. Fig. 4(b) shows the average SE performance for the three systems. Distributed sole-FD system shows the highest SE performance due to the fact that all antennas are deployed in a distributed manner and work in FD. However, our proposed system has only marginal SE loss compared to the distributed sole-FD, since the DAs far from users are turned off to significantly reduce power consumption. The co-located FD MIMO system demonstrates the poorest SE performance, which is because the co-located antenna deployment is sensitive to the distance between users and antennas and has limited ability in reducing PL. The co-located antenna deployment has also resulted in the strongest self-interference among different antennas. Therefore, the co-located FD MIMO system requires higher SIC amount than the DA deployment. As a result, the co-located FD MIMO system’s SE increases more slowly than the DA systems.

Fig. 5(a) and Fig. 5(b) show the effects of power constraint $p_{DA}$ on EE and power consumption, respectively. As shown in Theorem 1, EE is a quasi-concave function with respect to the transmission power. If the power constraint is relatively low, EE is mono-increasing. If the power constraint is relatively high, the global optimal EE can be achieved and the EE performance approaches saturation, as showed in Fig. 5(a). Also, it can be seen that the proposed system outperforms the others with all the transmission power constraints, even when transmission power is only 60 mW. It is because the proposed system can activate/deactivate DAs, and adjust the downlink beamformer and uplink transmission power to make a trade-off between SE and power consumption. Fig. 5(b) shows that the proposed algorithm enables the least power consumption among the three systems. The co-located FD MIMO consumes the highest power. This is because the incurred circuit power consumption in the co-located FD MIMO system plays a dominant role in the total power consumption, and all antennas are kept active to pursue high SE without significant increase in the total power consumption. As a result, the co-located FD MIMO prefers to fully utilize all the transmission power to pursue high SE.

Fig. 6 shows the effect of threshold value $\psi$ on power consumption and SE performances. $p_{DA}$ on EE and power consumption, respectively. As shown in Theorem 1, EE is a quasi-concave function with respect to the transmission power. If the power constraint is relatively low, EE is mono-increasing. If the power constraint is relatively high, the global optimal EE can be achieved and the EE performance approaches saturation, as showed in Fig. 5(a). Also, it can be seen that the proposed system outperforms the others with all the transmission power constraints, even when transmission power is only 60 mW. It is because the proposed system can activate/deactivate DAs, and adjust the downlink beamformer and uplink transmission power to make a trade-off between SE and power consumption. Fig. 5(b) shows that the proposed algorithm enables the least power consumption among the three systems. The co-located FD MIMO consumes the highest power. This is because the incurred circuit power consumption in the co-located FD MIMO system plays a dominant role in the total power consumption, and all antennas are kept active to pursue high SE without significant increase in the total power consumption. As a result, the co-located FD MIMO prefers to fully utilize all the transmission power to pursue high SE.

Fig. 6 shows the effect of threshold value $\psi$ on power consumption and SE performances, where only our proposed system is plotted. It can be seen that, less power and lower throughput are demonstrated with a higher threshold, since fewer DAs are activated. This indicates that we can adjust the threshold value to reduce the power consumption, at the cost of SE. On the other hand, a lower threshold allows a better SE performance, while power consumption is relatively higher, which corresponds to a stringent SE QoS scenario. This confirms our analysis in the DA clustering algorithm that by changing the threshold value $\psi$, the number of activated DAs can be well controlled towards different QoS requirements.

Fig. 7 shows the required number of self-interference reference chains by the three systems. Since the analog and digital domain cancellations need to tap reference signal from every transmit chain and feed it to all receive chains for SIC operation [1] [38], the required number of self-interference
reference chains has a significant impact on the SIC circuit design. It can be seen that, the proposed system requires the fewest self-interference reference chains. It is because DAs may be in sleep or HD mode, where SIC operation is not needed. Besides, benefiting from the distributed deployment, passive suppression in the propagation domain guarantees a reasonable SIC amount across different DAs, and thus the analog and digital domain SIC are not required across different DAs. For the DA sole-FD systems in [31], DAs can only work in either FD or sleep mode, so the number of reference chains is equivalent to the number of active DAs. For FD MIMO, since all antennas are active and work in FD mode [47] [52], the required number of reference chains is proportional to the number of antennas. As a result, the co-located FD MIMO system requires much more reference chains than others.

Fig. 8 shows the average probabilities of working modes of different DAs, where axes $x$ and $y$ represent distance as illustrated in Fig. 2. According to the DA clustering algorithm, only those DAs with good channel conditions in uplink or downlink are turned on for FD mode, and therefore the DAs in the central part of the area have a higher probability of working in FD mode. In contrast, the DAs located at cell edge have a higher probability of working in HD mode or sleeping.

Fig. 9 shows the effect of the number of users on the probabilities of working modes averaged across 25 DAs. With more users, the DAs have a higher probability of working in FD mode while the probabilities of HD and sleep modes degrade. According to the DA clustering algorithm, only the DAs with acceptable channel gain in two directions can work in FD mode. As a result, with more users, the distance between DAs and users is statistically shortened, enabling the increase in the probability of working in FD mode and decrease in the probabilities of HD and sleep modes.

VI. CONCLUSION

We have proposed a novel bi-directional DA system with hybrid duplexing, which features high EE, low power consumption, and simple SIC circuit design. Based on the proposed system, the EE-oriented resource allocation has been investigated by jointly designing activation/deactivation of transmit/receive chains, downlink beamformer and uplink transmission power. Simulation results show that the proposed design demonstrates significant EE enhancement and power saving over the co-located FD MIMO system and the DA sole-FD system, achieving a reasonable balance between low power consumption and high SE. Moreover, since DAs can flexibly adjust their duplexing and working modes due to the enhanced degree of freedom, the proposed system obtains a much simpler SIC circuit design compared to the two benchmark systems in [31] and [47].

APPENDIX A

PROOF OF THEOREM 1

Define the superlevel set of $\omega(W, p)$ as $S_{\delta} = \{W, p \in \Theta | \omega(W, p) \geq \delta \}$. According to [19], $\omega(W, p)$ is jointly quasi-
concave with respect to variables $W$ and $p$ if $S_\delta$ is convex for any real number $\delta$. When $\delta \leq 0$, there is no physical meaning. When $\delta \geq 0$, $S_\delta$ is equivalent to $S_\delta = \{ W, p \in \Theta | p_{total}(W, p) - \delta f_1(W, p) - f_2^n(W, p) \leq 0 \}$. According to our analysis, $p_{total}(W, p)$ is affine with respect to the variables, while $-\delta (f_1(W, p) - f_2^n(W, p))$ is strictly jointly convex with respect to the variables. Therefore, the summation is strictly convex with respect to the variables, and $\omega(W, p)$ is quasi-concave with respect to the variables.

**APPENDIX B**

**EQUIVALENCE BETWEEN MAXIMIZING PROBLEM P4 AND FINDING THE ROOT OF PROBLEM P5**

The proof is similar to and based on the Lemma 1 in [48]. Suppose that $(W^*, p^*)$ is the optimal solution of (23), with the optimal $\beta^*$. Define $\beta_0$ as the root of $f_1(W, p) - f_2^n(W, p) - \beta p_{total}(W, p) = 0$. If $f_1(W, p) - f_2^n(W, p) - \beta_0 p_{total}(W, p) < 0$ for all $(W, p) \in \Theta$, including $W^*, p^*$. Therefore, we have $\beta^* \leq \beta_0$. As a result, $\beta_0 = \beta_0$ is lead.

On the other hand, if $\beta = \beta^*$, at least we have $f_1(W^*, p^*) - f_2^n(W^*, p^*) - \beta_0 p_{total}(W^*, p^*) = 0$ and thus $f_1(W, p) - f_2^n(W, p) - \beta_0 p_{total}(W, p) \leq 0$ holds. As a result, the equality $\beta = \beta^*$ yields $f_1(W, p) - f_2^n(W, p) - \beta_0 p_{total}(W, p) = 0$. In conclusion, the sufficient and necessary condition of $\beta = \beta^*$ is $f_1(W, p) - f_2^n(W, p) - \beta_0 p_{total}(W, p) = 0$, and the equivalence between maximizing problem P4 and finding the root of P5 is proven.

**APPENDIX C**

**PROOF OF RANK ONE**

The SDP relaxed problem in (23) satisfies the Slater’s constraint qualification. To obtain the dual problem, the Lagrange function of the primal problem is given by

$$L = f_1(W, p) - f_2^n(W, p) - \beta p_{total}(W, p) + \sum_{l=1}^L \zeta_l \left( p_{DA} - \sum_{k=1}^K Tr(W_k F_l) \right) + \sum_{k=1}^K W_k Z_k + \sum_{u=1}^U \sum_{u=1}^U \sum_{u=1}^U \lambda_u p_u + \sum_{u=1}^U \sum_{u=1}^U \lambda_u p_u,$$

(25)

where $\lambda_u$, $\zeta_l$, $Z_k$ and $S_u$, for all $l \in L, u \in U, k \in K$, are Lagrange multipliers associated with the constraints in (23).

According to (25), the Karush-Kuhn-Tucker (KKT) conditions used for the proof include the dual constraints: $Z_k^* \succeq 0$ and $\zeta_l^* \geq 0$; complementary slackness: $W_k^* Z_k^* = 0$; and the gradient of Lagrange function (25) with respect to $W_k$ vanishing to 0: $\frac{\partial}{\partial W_k} L = 0$. In particular, $\frac{\partial L}{\partial W_k} = 0$ can be further detailed by (26)

$$Z_k^* = A_k^* - K \sum_{m=1}^L \sum_{n=1}^L t_m t_n F_m^T (F_n^H)^T H_k,$$

(26)

where variables $A_k^*$, $\Gamma^*$ and $\Lambda^*$ involved in (26) are given by (27) (28) and (29), respectively. By pre-multiplying equality (26) by $W_k^*$, we have

$$W_k^* A_k^* = W_k^* K \sum_{m=1}^L \sum_{n=1}^L t_m t_n F_m^T (F_n^H)^T H_k,$$

(30)

Since all channel variables in the system are statistically independent and the matrices span the whole signal space, we know that $A_k^*$ is a full-rank matrix [34]. According to the properties of matrix calculus, we have

$$\text{Rank}(W_k^*) = \text{Rank}(W_k^* A_k^*) = \text{min} \left\{ \text{Rank}(W_k^*), \text{Rank}(K \sum_{m=1}^L \sum_{n=1}^L t_m t_n F_m^T (F_n^H)^T H_k) \right\} \leq \text{Rank}(K \sum_{m=1}^L \sum_{n=1}^L t_m t_n F_m^T (F_n^H)^T H_k) \leq 1.$$

(31)

Now, we have the inequality of $\text{Rank}(W_k^*) \leq 1$. We note that only a null matrix has rank 0 while $W_k^* \neq 0$. Thus, $\text{Rank}(W_k^*) = 1$ is obtained.

**APPENDIX D**

**PROOF OF CONVERGENCE IN INNER LAYER BY THE FW METHOD**

Hereby we prove the convergence in the inner layer by the FW method [11]. Let $T^{(n+1)}$ denote the optimal value of (P5) at the $(n+1)$-th iteration. According to the updating method in FW, we have

$$T^{(n+1)} = f_1(W^{(n+1)}, p^{(n+1)}) - f_2^n(W^{(n+1)}, p^{(n+1)}) - \beta p_{total}(W^{(n+1)}, p^{(n+1)}) = \text{argmax}_{W, p} f_1(W, p) - f_2^n(W, p) - \beta p_{total}(W, p).$$

(32)

$T^{(n+1)}$ is the optimal value at the $(n+1)$-th iteration, which is greater than any other solutions in the feasible domain. Therefore, we have the following inequality that

$$\text{argmax}_{W, p} f_1(W, p) - f_2^n(W, p) - \beta p_{total}(W, p) \geq f_1(W^{(n)}, p^{(n)}) - f_2^n(W^{(n)}, p^{(n)}) - \beta p_{total}(W^{(n)}, p^{(n)}).$$
$A_k = \sum_{k' \neq K}^{K} \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \frac{F_m^T (F_n^H) T}{\alpha_k} H_{k'} + \frac{1}{\eta} I + \sum_{l=1}^{L} \zeta_l F_l +$

$$\sum_{u=1}^{U} \frac{1}{\alpha} \text{diag} \left( H_{SI} \left( \sum_{k=1}^{K} \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \frac{F_m^T (F_n^H) T}{\alpha_k} H_{SI}^H \right) \right) \left( \sum_{k=1}^{K} \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \frac{F_m^T (F_n^H) T}{\alpha_k} H_{SI}^H \right)$$

$$U \frac{1}{\alpha} \text{Tr} \left( \text{diag} \left( H_{SI} \left( \sum_{k=1}^{K} \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \frac{F_m^T (F_n^H) T}{\alpha_k} H_{SI}^H \right) \right) \right)$$

$$\Lambda^* = \sum_{k=1}^{K} \log_2 \left( \sum_{k'=1}^{K} \text{Tr} \left( \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n F_m W_{kn} F_n^H H_{kn} \right) + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2 \right).$$

$$\sum_{u=1}^{U} \frac{1}{\alpha} \text{Tr} \left( \text{diag} \left( H_{SI} \left( \sum_{k=1}^{K} \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n F_m W_{kn} F_n^H H_{kn} \right) \right) \right)$$

Since we have the equalities that $f_2^{(n)}(W^{(n)}, p^{(n)}) = f_2(W^{(n)}, p^{(n)})$ and $\nabla f_2^{(n)}(W^{(n)}, p^{(n)}) = \nabla f_2(W^{(n)}, p^{(n)})$, (33) can be further derived as

$$f_1(W^{(n)}, p^{(n)}) - f_2^{(n)}(W^{(n)}, p^{(n)}) = \beta P_{\text{total}}(W^{(n)}, p^{(n)})$$

$$f_1(W^{(n)}, p^{(n)}) - f_2(W^{(n)}, p^{(n)}) = \beta P_{\text{total}}(W^{(n)}, p^{(n)})$$

$$f_1(W^{(n)}, p^{(n)}) - f_2^{(n-1)}(W^{(n)}, p^{(n)}) - \beta P_{\text{total}}(W^{(n)}, p^{(n)}) = \nabla f_1(W^{(n)}, p^{(n)})$$

$$f_1(W^{(n)}, p^{(n)}) - f_2^{(n-1)}(W^{(n)}, p^{(n)}) = \nabla f_1(W^{(n)}, p^{(n)})$$

$$f_1(W^{(n)}, p^{(n)}) = \nabla f_1(W^{(n)}, p^{(n)})$$

$$f_1(W^{(n)}, p^{(n)}) = \nabla f_1(W^{(n)}, p^{(n)})$$

$$\nabla_{W_k} f_1(W^{(n)}, p^{(n)}) - \nabla_{W_k} f_2^{(n)}(W^{(n)}, p^{(n)}) + Z_k$$

$$- \nabla_{W_k} \beta P_{\text{total}}(W^{(n)}, p^{(n)}) - \sum_{l=1}^{L} \zeta_l F_l = 0,$$  \hspace{1cm} (35)

$$\nabla_{p_u} f_1(W^{(n)}, p^{(n)}) - \nabla_{p_u} f_2^{(n)}(W^{(n)}, p^{(n)}) - \zeta_u + \lambda_u - \nabla_{p_u} \beta P_{\text{total}}(W^{(n)}, p^{(n)}) = 0,$$  \hspace{1cm} (36)

$$\zeta_l \left( p_{DA} - \sum_{k=1}^{K} \text{Tr}(W_k^T F_l) \right) = 0,$$  \hspace{1cm} (37)

For all $k \in K, \forall u \in U$, and all $l \in L$. According to the equality that $\nabla f_2^{(n)}(W^{(n)}, p^{(n)}) = \nabla f_2(W^{(n)}, p^{(n)})$, we can replace $\nabla_{W_k} f_2^{(n)}(W^{(n)}, p^{(n)})$ and $\nabla_{p_u} f_2^{(n)}(W^{(n)}, p^{(n)})$ by $\nabla_{W_k} f_2(W^{(n)}, p^{(n)})$ and $\nabla_{p_u} f_2(W^{(n)}, p^{(n)})$, which are exactly the KKT conditions of the original problem $f_1(W, p) - f_2(W, p) - P_{\text{total}}(W, p)$. As a result, the FW method leads to a convergence in the inner layer.

**REFERENCES**


Zhongxiang Wei (S’15–M’17) received the Ph.D. degree in Electrical Engineering and Electronics from the University of Liverpool, Liverpool, U.K., in 2017. From Mar. 2016 to Mar. 2017, he was with the Institution for Infocomm Research (I2R), Singapore, as a Research Assistant. His research interests include green communications, full-duplex, millimeter-wave communications, resource allocation, and algorithm design. He received graduate China National Scholarship Award in 2012 and CSC Outstanding Self-Financed Scholarship in 2017.

Sumei Sun (F’16) is currently the Head of the Communications and Networks Cluster, Institute for Infocomm Research, Agency for Science, Technology, and Research, Singapore, focusing on smart communications and networks for robust, QoS/QoE-guaranteed, and energy- and spectrum-efficient connectivity for human, machine, and things. She publishes actively in IEEE journals and conferences, and collaborates closely in research with the industry. She is a Distinguished Lecturer of the IEEE Vehicular Technology Society 2014–2018, a Distinguished Visiting Fellow of the Royal Academy of Engineering, U.K., in 2014, and Vice Director of IEEE Communications Society Asia Pacific Board since 2016. She has also been actively contributing to organizing IEEE conferences in different roles, including her recent services as Executive Vice Chair of Globecom 2017, Symposium Co-Chair of ICC 2015 and 2016, Track Co-Chair of IEEE VTC 2016 Fall, VTC 2017 Spring, etc. She was an Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (TVT) during 2011–2017, and Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS during 2011–2016. She serves as an Area Editor of IEEE TVT since July 2017, and Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, since 2015. She received the Top Associate Editor award in 2011, 2012, and 2015, all from IEEE Transactions on Vehicular Technology.

Xu Zhu (S’02–M’03–SM’12) received the BEng degree (with the first class honors) in Electronics and Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 1999, and the PhD degree in Electrical and Electronic Engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2003. Since May 2003, she has been with the Department of Electrical Engineering and Electronics, the University of Liverpool, Liverpool, UK, where she is currently a Reader. Dr. Zhu has over 150 peer-reviewed publications on communications and signal processing. She has served as an Editor for the IEEE Transactions on Wireless Communications from 2012–2017, and a Guest Editor for a number of international journals such as Electronics. She has organized various international conferences as Chair, such as Symposium Co-Chair of the IEEE ICC 2016 and 2019, Vice Chair of the 2006 and 2008 ICARN International Workshops, Program Chair of the ICSAI 2012, and Publicity Chair of the IEEE IUCC-2016. Her research interests include MIMO, channel equalization, resource allocation, optimization, cooperative communications, green communications etc..

Yi Huang (S’91–M’96–SM’06) received the DPhil in Communications from the University of Oxford, UK in 1994. He has been conducting research in wireless communications, applied electromagnetics, radar and antennas since 1987. His experience includes 3 years spent with NRIET (China) as a Radar Engineer and various periods with the Universities of Birmingham, Oxford, and Essex at the UK. He worked as a Research Fellow at British Telecom Labs in 1994, and then joined the University of Liverpool, UK in 1995, where he is now a full Professor and Deputy Head of the Department of Electrical Engineering and Electronics. Dr. Huang has published over 300 refereed papers, received many research grants from various funding bodies and acted as consultant for various companies. He has been a keynote/invited speaker and organizer of many conferences and workshops (e.g., WiCom2010, IEEE iWAT 2010 and LAPC2012). He is the Editor-in-Chief of Wireless Engineering and Technology and has been an Editor or Guest Editor for four international journals. He is a UK national representative of EU COSTIC1102 and Executive Committee Member and Fellow of the IET.

Jingjing Wang (S’14) received his B.S. degree in Electronic Information Engineering from Dalian University of Technology in 2014 with the highest honors. He currently works for his PhD degree at Complex Engineered Systems Lab (CESL) in Tsinghua University, Beijing. From Jan. 2016 to Mar. 2016, he visited Wireless Networks and Decision Systems (WNDS) Group, Singapore University of Technology and Design as a visiting student. Since Oct. 2017, he has been visiting the Telecommunications Group, University of Southampton as a joint PhD student. His research interests include the resource allocation and network association, learning theory aided modeling, analysis and signal processing, and information diffusion theory. He received Tsinghua GuangHua Scholarship in 2016 and graduate China National Scholarship Award in 2017.